Intermediate Consistencies by Delaying Expensive Propagators

Andreï Legtchenko, Arnaud Lallouet, AbdelAli Ed-Dbali
Université d’Orléans – LIFO
BP 6759 – F-45067 Orléans – France

Abstract
What makes a good consistency? Depending on the constraint, it may be a good pruning power or a low computational cost. By “weakening” arc-consistency, we propose to define new automatically generated solvers which form a sequence of consistencies weaker than arc-consistency. The method presented in this paper exploits a form of regularity in the cloud of constraint solutions: the density of solution orthogonal to a projection. This approach is illustrated on the sparse constraints “to be a n-letters english word” and crossword CSPs, where interesting speed-ups are obtained.

Introduction
Since their introduction (Mackworth 1977), CSP consistencies have been recognized as one of the most powerful tool to strengthen search mechanisms. Since then, their considerable pruning power has motivated a lot of efforts to find new consistencies and to improve the algorithms to compute them.

Consistencies can be partially ordered according to their pruning power. However, this pruning power should be put into balance with the complexity of enforcing them. For example, path-consistency is often not worth it: its pruning power is great, but the price to pay is high. Maintaining path-consistency during search is thus often beaten in practice by weaker consistencies. Similarly, on many useful CSPs, bound-consistency is faster than arc-consistency even if it does not remove values inside the variables’ domains: this is left to the search mechanism ensuring the completeness of constraint solving. Moreover, the overall performance of a solver is also determined by the interaction between the consistency and the search strategy. Powerful consistencies are not always the best choice. It is sometimes more interesting to find the optimal ratio between the advantage of pruning and the computational cost needed to enforce it. The absence of prediction of the potential power of a consistency comes from two sources: consistencies exploit subtle properties which are often incomparable and they do not always use the same data-structures.

Recently, there has been a great interest in the automatic building of consistency operators and in finding ad-hoc efficient constraint representations. First (Apt & Monfroy 2001) and (Abdennadher & Rigotti 2003) use CHRs (Frühwirth 1998) as a target language to build a solver. Powerful propagators may be discovered in this framework but at the price of a high complexity which limits the method to constraints of modest arities and rather small domains. In (Dao et al. 2002) has been introduced an approximation framework we also use in this paper to approximate bound-consistency. This work has been extended in (Lallouet et al. 2003) for arc-consistency: a constraint is approximated by a set of covering blocks obtained by a clustering algorithm and only these blocks are used in the reduction process. But unfortunately, not all constraints are suitable for clustering. It is needed that the constraint has locally dense areas which can be agglomerated meaningfully. Finding new representations to compute a consistency has also been tackled in (Barták & Mecl 2003) and (Cheng, Lee, & Stuckey 2003): they use clustering techniques to find an exact representation of the constraint. Hence the resulting consistency is the full arc-consistency and speed-ups are difficult to obtain while competing with highly optimized implementations. In (Monfroy 1999), the idea of using weaker functions has been proposed as a preprocessing to speed-up the computation of a classical consistency (in this case, box-consistency). From the search mechanism point of view, this is an internal recipe to compute the consistency. In contrast, our operators define on purpose a weaker consistency which is used during search.

In this paper, we contribute to this line of work by providing a range of custom consistencies intermediate between bound- and arc-consistency for sparse constraints which have no dense areas. First, these constraints exist: we take as example the n-ary constraint “to be a n-letter english word”. There is absolutely no block regularity and the words are sparsely distributed in the solution space. For example, by using the 4176 five-letters words of /usr/dict/words, we get a density of 0.035%. A clustering algorithm would either consider a large number of clusters, limiting the interest of the representation, or filter a too few number of non-solution out of the search space, yielding a very weak consistency.

Our method is as follows. For a given constraint, we express the arc-consistency as a set of particular elementary reduction functions. Each elementary function is only con-
cerned in the withdrawal of one value in a variable’s domain. Depending on the number of supports for the target value, these functions do not have the same computational cost. Our technique consists in applying the computationally cheap operators as a consistency during the reduction phase and delaying the expensive ones until instantiation of all variables. For this we split the set of functions in two: functions smaller than a given threshold are iterated and longer ones are delayed. The closure of all the operators defines a new consistency for a given constraint, weaker and sometimes quicker than arc-consistency.

Intuitively, the cost of an elementary reduction is directly related to the number of supports present in the hyperplane orthogonal to the target value. For example, in figure 1 representing cuts in the solution space of a constraint, the value \( a \) for the variable \( X \) is supported by a large number of tuples: in order to eliminate \( a \) from \( X \)’s domain, we have to check all these tuples. On the other side, the value \( b \) in figure 2 is only supported by a few tuples. But since \( a \) is well supported, it is more likely part of a solution than \( b \). Hence we spend a large work trying to eliminate a value which is likely part of a solution. We propose to postpone the evaluation of \( a \)’s function until instantiation of all variables (i.e. on satisfiability test) while actively trying to eliminate \( b \) by consistency. It could be argued that algorithms like AC6 which compute lazily the set of supports could be faster, but since the resulting consistency is weaker than AC, the technique could apply to any AC algorithm. The main difficulty rely in determining the best threshold.

The paper is structured as follows:

- A framework to express consistencies. Consistencies are usually built as global CSP properties. But it is now rather common to express them modularly by the greatest fix-point of a set of operators associated to the constraints, which is computed using a chaotic iteration (Apt 1999). We present a framework which allows, starting from arbitrary operators, to progressively add properties in order to build a consistency.

- A consistency construction method. For a given constraint, we express the arc-consistency as a set of particular reduction operators. These operators do not have the same computational cost. Expensive ones are delayed until instantiation of all variables. The closure of all the operators defines a new consistency for a given constraint, weaker but quicker than arc-consistency.

- An example. The interest of these consistencies is shown by the ”word” constraints and the crossword CPSs.

An approximation framework

In this section, we present a general approximation scheme for consistencies. Let \( V \) be a set of variables and \( D = \{D_X\}_{X \in V} \) their (finite) domains. For \( W \subseteq V \), we denote by \( D^W \) the set of tuples on \( W \), namely \( \Pi_{X \in W} D_X \). Projection of a tuple or a set of tuples on a variable or a set of variables is denoted by \(|\cdot|\), natural join of two sets of tuples is denoted by \( \ltimes \). If \( A \) is a set, then \( \mathcal{P}(A) \) denotes its powerset and \( |A| \) its cardinal.

**Definition 1 (Constraint)** A constraint \( c \) is a pair \((W,T)\) where:

- \( W \subseteq V \) is the arity of the constraint \( c \) and is denoted by \( \text{var}(c) \).
- \( T \subseteq D^W \) is the set of solutions of \( c \) and is denoted by \( \text{sol}(c) \).

The join of two constraints is defined as a natural extension of the join of tuples: \( c \ltimes c' = (\text{var}(c) \cup \text{var}(c'), \text{sol}(c) \ltimes \text{sol}(c')) \).

A CSP is a set of constraints. Join is naturally extended to CSPs and the solutions of a CSP \( C \) are \( \text{sol}(\ltimes C) \). A direct computation of this join is too expensive to be tractable, especially when considering that it needs to represent tuples of the CSP’s arity. This is why a framework based on approximations is preferred, the most successful of them being the domain reduction scheme where variable domains are the only reduced constraints. So, for \( W \subseteq V \), a search state consists in a set of yet possible values for each variable: \( s_W = \{s_X\}_{X \in W} \) such that \( s_X \subseteq D_X \). The search space is \( S_W = \Pi_{X \in W} \mathcal{P}(D_X) \). The set \( S_W \), ordered by pointwise inclusion \( \subseteq \), is a complete lattice. Likewise, union and intersection on search states are defined pointwise. The whole search space \( S_V \) is simply denoted by \( S \).

Some search states we call singletonic play a special role in our framework. A singletonic search state comprises a single value for each variable, and hence represents a single
A tuple is promoted to a singleton search state by the operator \([\cdot]_T\): for \(t \in D^W\), let \([t]_T = \{(t_X)_{X \in W}\} \subseteq S_W\). This notation is extended to a set of tuples: for \(E \subseteq D^W\), let \([E]_T = \{[t]_T \mid t \in E\} \subseteq S_W\). Conversely, a search state is converted into the set of tuples it represents by taking its cartesian product \(\Pi\): for \(s \in S_W\), \(\Pi s = \Pi X \in W s_X \subseteq D^W\). We denote by \(Sing_W\) the set \([D^W]\) of singleton search states. By definition, \([D^W]\) \(\subseteq S_W\).

A consistency is generally described by a property \(Cons \subseteq S\) which holds for certain search states and is classically modeled by the common greatest fixpoint of a set of operators associated to the constraints. By extension, in this paper, we call consistency for a candidate tuple can be done by the propagation mechanism itself. Let \(C = \{c_1, \ldots, c_n\}\) be a CSP and \(F = \{f_1, \ldots, f_n\}\) be a set of consistencies on \(S\) associated respectively to \(\{c_1, \ldots, c_n\}\). If all constraints are not defined on the same set of variables, it is always possible to use the extension of the operators on the union of all variables which appear in the constraints. The common closure of the operators of \(F\) can be computed by a chaotic iteration (Apt 1999). It follows from the main confluence theorem of chaotic iterations that a consistency can be constituted by combining the mutual strengths of several operators. Since we have \(Fix(F) = \bigcap_{f \in F} Fix(f)\) and since each consistency does preserve the tuples of its associated constraint, the computed closure of all operators associated to the CSP \(C\) does not reject a tuple of \(c_1 \times \ldots \times c_n\) for any search state \(s \in S\) because of an operator of \(F\).

**Proposition 5 (Composition)** The composition of two preconsistencies for a constraint \(c\) via chaotic iteration is still a preconsistency for \(c\).

The proof is straightforward by (Apt 1999). We call \(\odot\) the composition of two operators via chaotic iteration. The same property holds for consistencies instead of preconsistencies.
The goal of an automatic construction procedure is to build for a constraint a set of operators which can ensure a desired level of consistency when included in a chaotic iteration. We can compare two operators \( f_1 \) and \( f_2 \) by their reductions over the search space:

**Definition 7** An operator \( f_1 \) is stronger than \( f_2 \), denoted by \( f_1 \subseteq f_2 \) if \( \forall s \in S, f_1(s) \subseteq f_2(s) \).

For example, it is well-known that arc-consistency is stronger than bound-consistency which in turn is stronger than \( id_c \). The notion of preconsistency is interesting in the context of CSP resolution because preconsistency operators have the desired properties to be included in a chaotic iteration. Since they are weaker than consistencies, they are easier to construct automatically. But since they may accept some non-solution singleton states, it is needed to find a technique to ensure completeness. The framework we propose is to build a fast but incomplete preconsistency and to add *repair* operators which are delayed until the satisfiability test.

**Delaying “long” operators**

In order to build operators for an intermediate consistency, we start by giving the expression of arc-consistency with a set of special function we call *elementary reduction functions*. Then we weaken the arc-consistency by delaying a set of functions which are supposed to be too expensive until instantiation of all variables. We use as a measure of the syntactic expression of the function. Hence the reduction power decreases, but the computation becomes quicker. When the choice of the length threshold varies, it defines a sequence of comparable consistencies.

**Definition 8 (Support)** Let \( c = (W, T) \) be a constraint, \( X \subseteq W \) and \( a \in D_X \). We call support of “\( X = a \)” a tuple \( t \in \text{sol}(c) \) such that \( t_X = a \).

**Example 9** Let \( c(X, Y, Z) \) be a constraint. \( D_X = D_Y = \{0, 1\} \) and \( D_Z = \{0, 1, 2\} \).

\[
\begin{array}{ccc}
X & Y & Z \\
0 & 0 & 0 \\
0 & 1 & 0 \\
1 & 0 & 0 \\
1 & 1 & 1 \\
1 & 1 & 2 \\
0 & 1 & 2 \\
\end{array}
\]

The sets of supports for \( Z \) are:

\[
\begin{align*}
T_{Z=0} &= \{(0, 0, 0), (0, 1, 0), (1, 0, 0)\} \\
T_{Z=1} &= \{(1, 1, 1)\} \\
T_{Z=2} &= \{(1, 1, 2), (0, 1, 2)\}
\end{align*}
\]

We call \( T_{X=a} \subseteq \text{sol}(c) \) the set of all supports of \( X = a \). A value \( a \) has to be maintained in the current domain of \( X \) only if we have at least one \( t \in T_{X=a} \) such that all projections on \( Y \subseteq W \setminus \{X\} \) are included in \( s_Y \).

**Definition 10 (Supp property)** Let \( c = (W, T) \) be a constraint, \( X \subseteq W \) and \( a \in D_X \). We call \( \text{Supp}_{X=a}(s) \) the following property:

\[
\bigvee_{t \in T_{X=a}} \left( \bigwedge_{Y \in W \setminus \{X\}} t_Y \in s_Y \right)
\]

The value \( X = a \) is supported if \( \text{Supp}_{X=a}(s) = \text{true} \).

**Example 11 (Example 9 running)** For all \( s \in S_{(X,Y,Z)} \), we have:

\[
\begin{align*}
\text{Supp}_{Z=0}(s) &= (0 \in s_X \land 0 \in s_Y) \lor (0 \in s_X \land 1 \in s_Y) \\
&\lor (1 \in s_X \land 0 \in s_Y) \\
\text{Supp}_{Z=1}(s) &= 1 \in s_X \land 1 \in s_Y \\
\text{Supp}_{Z=2}(s) &= (1 \in s_X \land 1 \in s_Y) \lor (0 \in s_X \land 1 \in s_Y).
\end{align*}
\]

If \( X = a \) is not supported, then a does not participate to any solution of the CSP and therefore can be eliminated. With this notion, we define a function called *elementary reduction function* which concerns only one value in the variable’s domain. Each value in the initial domain of each variable has its own elementary reduction function. If a value must be eliminated, its function returns this value as a singleton, and the empty set otherwise.

**Definition 12 (Elementary reduction function)** For all \( X \subseteq W \) and for all \( a \in D_X \), the elementary reduction function associated to \( X = a \) is \( r_{X=a} : S_W \longrightarrow \mathcal{P}(D_X) \) given by

\[
\forall s \in S_W, \quad r_{X=a}(s) = \begin{cases} \{a\}, & \text{if } \neg \text{Supp}_{X=a}(s) \\ \emptyset, & \text{otherwise} \end{cases}
\]

We call \( s\)-size (support size) of the elementary reduction function \( r_{X=a} \) the cardinality of \( T_{X=a} \). Arc-consistency can be defined using elementary reduction functions as follows:

\[
\forall s \in S_W, \quad ac_s(s) = (s_X \cup \{a\} \cup \{b\} \in D_X, r_{X=a}(s)) x \in W.
\]

We want a consistency quiker than arc-consistency, even if it is less powerful. For this, we choose a threshold and split arc-consistency into two operators. The first one, called *Short*, is composed of all elementary reduction functions having their \( s\)-size less than the chosen threshold.

**Definition 13 (Operator Short)** Let \( c \) be a constraint, and \( n \) an integer. The operator \( \text{Short}_c(n) : S_W \longrightarrow S_W \) is given by:

\[
\forall s \in S_W, \quad \text{Short}_c(n)(s) = \left( s_X \cup \{a\} \cup \{b\} \in D_X, r_{X=a}(s) \right) x \in W.
\]

The second operator, called *Long*, is composed of all other elementary reduction functions, which are fired only on singleton states for completeness.

**Definition 14 (Operator Long)** Let \( c \) be a constraint and \( n \) an integer. The operator \( \text{Long}_c(n) : S_W \longrightarrow S_W \) is given by:

\[
\forall s \in S_W, \quad \text{Long}_c(n)(s) = \begin{cases} \{s_X \cup \{a\} \cup \{b\} \in D_X, r_{X=a}(s) \} x \in W, & \text{if } s \in \text{Sing}_W \\ s, & \text{otherwise} \end{cases}
\]

This operator is used only to reject non-solution tuples.

**Example 15 (Example 11 running)** For the variable \( Z \), we can make two operators following the definition of Short, and Long, Let set the threshold to 2. The operators for \( Z \) are:

\[
\begin{align*}
\text{Short}_Z(2)_Z: & \quad s_Z = s_Z \setminus (r_{Z=1}(s) \cup r_{Z=2}(s)) \\
\text{Long}_Z(2)_Z: & \quad s_Z = s_Z \setminus r_{Z=0}(s)
\end{align*}
\]

The operator \( \text{Long}_Z(2)_Z \) is delayed because its reduction power is small (only one value can be eliminated), and the effort to compute \( r_{Z=0}(s) \) is judged too high. This operator is fired only if \( s \in \text{Sing}_{X,Y,Z} \).
Proposition 16  For all integer $n$, Short$_c(n)$ is a preconsistency for $c$.

Proof  First, we show that Short$_c(n)$ is monotonic. Let $s$ and $s'$ in $S_W$ such that $s \subseteq s'$. Then, $\forall X \in W, \forall a \in D_X$, Supp$_{X=a}(s) \Rightarrow$ Supp$_{X=a}(s')$. From which it follows that $\forall X \in W, \forall a \in D_X, \neg$Supp$_{X=a}(s') \Rightarrow \neg$Supp$_{X=a}(s)$. In the case of $s'$, there are less values to eliminate than in the case of $s$. So Short$_c(n)(s) \subseteq$ Short$_c(n)(s')$. The operator Short$_c(n)$ is monotonic.

Short$_c(n)$ is contracting by construction. It it also correct by construction: it eliminates less values than arc-consistency. Hence Short$_c(n)$ is a preconsistency. □

The operator Long$_c(n)$ is also a preconsistency, the proof is similar to Short$_c(n)$. When both operators are in a chaotic iteration, we get a consistency:

Proposition 17  Let $c = (W, T)$ a constraint and $n$ an integer. The composition Short$_c(n) \circ$ Long$_c(n)$ is a consistency for $c$.

Proof  According to the proposition 5, Short$_c(n) \circ$ Long$_c(n)$ is a preconsistency. But $\forall s \in Sing_W$, Short$_c(n) \circ$ Long$_c(n)(s) = ac_c(s)$, so Short$_c(n) \circ$ Long$_c(n)$ is associated to $c$. Therefore Short$_c(n) \circ$ Long$_c(n)$ is a consistency for $c$. □

Elementary reduction functions of Long$_c(n)$ are not fired on non-singletonic states. If the threshold $n$ is not too big, the set of elementary reduction functions of Long$_c(n)$ is not empty. In that case, the consistency Short$_c(n) \circ$ Long$_c(n)$ is weaker than arc-consistency, but it can be computed faster, since that we have less functions to evaluate. If $n$ is large enough, all elementary reduction functions are in Short$_c(n)$ and we get the same reduction power and computational cost as arc-consistency.

Implementation and Example

Implementation.  A system generating the operators Short$_c(n)$ and Long$_c(n)$ has been implemented. The language used to express the operators is the indexical language (van Hentenryck, Saraswat, & Deville 1991) of GNU-Prolog (Diaz & Codognet 2001). An indexical operator is written "X in r" where X is the name of a variable, and r is a range of possible values for X and which may depend on other variables’ current domains. If we call x the current domain of X, the indexical X in x can be read declaratively as the second-order constraint $x \subseteq r$ or operationally as the operator $x \mapsto x \cap r$. For a given constraint and a threshold, our system returns two sets of $|W|$ indexical operators, i.e. one for each variable. The first set defines the Long$_c(n)$ operator, and the second Short$_c(n)$. In total, we have $2 \ast |W|$ indexical operators for a constraint $c = (W, T)$. The closure by chaotic iteration of all $2 \ast |W|$ operators is equivalent to the consistency Short$_c(n) \circ$ Long$_c(n)$. The indexical operators for Long$_c(n)$ are delayed with a special trigger val in the indexical language which delays the operator until a given variable is instantiated. This is how we get that these operators are not iterated on non-singletonic search states. This delay is obtained at no cost because Gnu-Prolog uses separate queues. In all cases, the generation time is about one second.

Example.  The sparse distribution described in figure 1 and 2 occurs, for example, in the case of the constraints word$_3(X_1, X_2, \ldots, X_n)$. For $n = 3$, the constraint word$_3(X, Y, Z)$ means that XYZ is a 3-letters English word. We consider that domains are ordered by lexicographic ordering. In figure 4 is presented the projections of word$_3(X, Y, Z)$ on different planes. When using UNIX dictionary /usr/dict/words, this constraint has 576 solutions. The constraints word$_4(X, Y, Z, U)$ (with 2236 solutions) and word$_5(X, Y, Z, U, V)$ (with 4176 solutions) have the same regularity.

The CSP we use consists in finding a solution for crossword grids of different sizes. The CSP is composed only by a set of constraints word$_n$. The domain of all variables is $\{a, \ldots, z\}$. An example of a 7x7 grid and its model are presented table 1.

Table 1: A model for 7x7 grid

<table>
<thead>
<tr>
<th>X1_1</th>
<th>X1_2</th>
<th>X1_3</th>
<th>X1_4</th>
<th>X1_5</th>
<th>X1_6</th>
<th>X1_7</th>
</tr>
</thead>
<tbody>
<tr>
<td>X2_1</td>
<td>X2_2</td>
<td>X2_3</td>
<td>X2_4</td>
<td>X2_5</td>
<td>X2_6</td>
<td>X2_7</td>
</tr>
<tr>
<td>X3_1</td>
<td>X3_2</td>
<td>X3_3</td>
<td>X3_4</td>
<td>X3_5</td>
<td>X3_6</td>
<td>X3_7</td>
</tr>
<tr>
<td>X4_1</td>
<td>X4_2</td>
<td>X4_3</td>
<td>X4_4</td>
<td>X4_5</td>
<td>X4_6</td>
<td>X4_7</td>
</tr>
<tr>
<td>X5_1</td>
<td>X5_2</td>
<td>X5_3</td>
<td>X5_4</td>
<td>X5_5</td>
<td>X5_6</td>
<td>X5_7</td>
</tr>
<tr>
<td>X6_1</td>
<td>X6_2</td>
<td>X6_3</td>
<td>X6_4</td>
<td>X6_5</td>
<td>X6_6</td>
<td>X6_7</td>
</tr>
<tr>
<td>X7_1</td>
<td>X7_2</td>
<td>X7_3</td>
<td>X7_4</td>
<td>X7_5</td>
<td>X7_6</td>
<td>X7_7</td>
</tr>
</tbody>
</table>

Figure 4: Projections of word$_3(X, Y, Z)$ on the planes XY, YZ, XZ.
In this paper, we propose a new method which allows to build a full range of consistencies weaker but quicker than arc-consistency. The interest of this method is illustrated on crossword CSPs. Further work can be to consider other representations for other regularities. For every representation, there should exist suitable constraints and other for which it is inadequate. It also includes a way to determine dynamically the parameters at run-time in order to exploit all parts of the constraint optimally.

### References


